

Model Explainer

Šimon Jukl

2026-01-16

Notace

y = perceived duration (response)

L = interval length (seconds)

T = tempo (beats per second)

G = group (slow/medium/fast)

β = regression coefficients

u = subject-level random effects

Model 1 (raw response)

$$y \sim \mathcal{N}(\mu, \sigma)$$
$$\mu = (\beta_{\mu 0} + u_{\mu 0}) + (\beta_{\mu L} + u_{\mu L}) \cdot L + (\beta_{\mu T} + u_{\mu T}) \cdot T + \beta_{\mu TL} \cdot T \cdot L$$
$$\log \sigma = \beta_{\sigma 0} + \beta_{\sigma L} \cdot L$$

```
brm(  
  formula = formula = bf(  
    response ~  
      tempo +  
      interval_length +  
      tempo:interval_length +  
      (tempo + interval_length || participant),  
    sigma ~ interval_length  
)  
,  
  family = gaussian()  
)
```

Model 2 (ratio)

$$\frac{y}{L} \sim \mathcal{N}(\mu, \sigma)$$
$$\mu = (\beta_{\mu 0} + u_{\mu 0}) + (\beta_{\mu T} + u_{\mu T}) \cdot T$$
$$y = L [(\beta_{\mu 0} + u_{\mu 0}) + (\beta_{\mu T} + u_{\mu T}) \cdot T]$$
$$y = (\beta_{\mu 0} + u_{\mu 0}) \cdot L + (\beta_{\mu T} + u_{\mu T}) \cdot T \cdot L$$
$$\log \sigma = \beta_{\sigma 0} + \beta_{\sigma L} \cdot L$$

```
brm(  
  formula = bf(  
    response / interval_length ~  
      tempo +  
      (tempo || participant)  
)  
,  
  family = gaussian()  
)
```

Model 3 (log)

$$\begin{aligned}\log y &\sim \mathcal{N}(\mu_{ij}, \sigma) \\ \mu &= (\beta_{\mu 0} + u_{\mu 0}) + (\beta_{\mu L} + u_{\mu L}) \cdot L + (\beta_{\mu T} + u_{\mu T}) \cdot T + \beta_{\mu TL} \cdot T \cdot L \\ y &= e^{(\beta_{\mu 0} + u_{\mu 0})} \cdot e^{(\beta_{\mu L} + u_{\mu L}) \cdot L} \cdot e^{(\beta_{\mu T} + u_{\mu T}) \cdot T} \cdot e^{\beta_{\mu TL} \cdot T \cdot L} \\ \log \sigma &= \beta_{\sigma 0} + \beta_{\sigma L} \cdot L\end{aligned}$$

```
brm(  
  formula = bf(  
    response_log ~  
      tempo +  
      interval_length +  
      tempo:interval_length +  
      (tempo + interval_length || participant),  
    sigma ~ interval_length  
)  
,  
  family = gaussian()  
)
```

Popis

Výše jsou specifikovány tři bayesovské modely pro analýzu dat. K jejich specifikaci byla použita package brms od Paula-Christiana Bürknera. Níže následuje vysvětlení jednotlivých proměnných, parametrů a dalších rozhodnutí v rámci výběru modelů.

Modely budou pomocí relevantních kritérií porovnány a bude vybrát model, který bude vykazovat nejlepší vlastnosti. (Bayesian workflow, Gelman et al., 2020)

Parametry

- **response** je odpověď participanta (délka držení mezerníku). Jde o odhad délky poslechnutého intervalu o určité délce, během kterého je slyšet tempo o určité frekvenci (Hz, bps).
 - V případě logaritmického modelu se jedná o logaritmus z response, musejí být všechny koeficienty modelu interpretovány nikoliv aditivně, ale multiplikativně (koeficienty jsou zlogaritmované, pro interpretaci potřeba exponovat $-> e^\beta$). Koeficienty pak značí procentuální změnu v response.

- `tempo` je proměnná označující frekvenci. Efekt lze interpretovat jako změnu v odpovědi za každý 1 Hz.
- `interval_length` je délka intervalu. Koeficient značí, že při změně o 1 s v délce intervalu se změní odpověď o hodnotu koeficientu.
- `tempo:interval_length` je interakce, která říká, že efekt délky intervalu je různý pro různá tempa. Tedy lidé budou odhadovat délku jinak, pokud jim během intervalu bude hrát tempo o různých frekvencích.
- `group` je hlavní efekt skupiny. Interpretace tohoto efektu je zavádějící, protože zde skupina vysoce koreluje s tempem. Kvůli značné kolinearitě s tempem navrhuji proměnou nezařadit.
- `tempo:group` je interakce zkoumající, zda je efekt tempa jiný v závislosti na skupině. Tato interakce je definována jen na dvou překrytých mezi skupinami (pro ostatní kombinace tempa a skupiny nejsou z definice našeho experimentu data, ze kterých by model mohl vycházet) a není tedy jasné, jak ji interpretovat. Není tedy též zařazena do modelů.
- `(1 + tempo + interval_length || participant)` - je očekáváno, že efekt tempa a délky intervalu bude odlišný pro různé lidi, jelikož každý vnímáme délku trochu jinak a také můžeme mít jiné "vnitřní počítadlo". Dvě čáry znamenají nekorelované per-subject parametry.

Priors

Priors budu nastaveny tak, aby u každého modelu při prior predictive checks vycházely realistické hodnoty pro response. (Tedy např. ne 100 vteřin nebo záporné hodnoty. Prostě realistické hodnoty naměřené délky pro participanty, kteří slyšeli nastavené délky intervalu.) Model bude samozřejmě podroben analýze citlivosti na specifikaci priorů. Při dostatečně velkém vzorku není očekáván jejich velký vliv na závěrečný model a ve výsledku by měl převážit vliv dat.